Chemical Biology (& Biological Chemistry)

Ben Davis [Ben.Davis@chem.ox.ac.uk] – 4 Lectures as Part of Option II

Prior Knowledge Required

1. *1st year Biological Chemistry course*: Molecules of life. What are amino acids, peptides, proteins, sugars, nucleotides, nucleic acids – knowledge of their structures. Energy & Phosphates. Protein structure and folding. Basic principles of enzyme catalysis and inhibitor action; Michaelis-Menten kinetics. Basics of protein structure. Genetics and Protein Synthesis; transcription/translation/DNA→mRNA→protein; one letter and three letter codes for amino acids.
2. *2nd year Biological Chemistry course*: the mechanisms and principles behind primary metabolism; nature's reagents; glycolysis; CAC; oxidative phosphorylation; amino acid metablolism.

There will be one 'foundation' class (**F**) revising some of the key aspects of your 1st and 2nd year lectures as part of this course: Monday of 5th week, 5-6pm, Wolfson Seminar Room, CRL.

Books:

"Structure and Mechanism in Protein Science" Fersht; Freeman 2000 *Good on Enzyme Basics & Folding*
"An Introduction to Biotransformations in Organic Chemistry" Hanson; *A good intro to biotrans*
"Biotransformations in Organic Chemistry" Faber; *More detailed aspects of biotrans*
"Biochemistry and Molecular Biology" Elliott & Elliott; OUP *Good for quick overview & basics*
"Foundations of Chemical Biology" OCP 98 *Good as a basic introduction (e.g. folding)*
"Carbohydrate Chemistry" OCP 99 *A Masterpiece – clearly.*
"Bioinformatics" Higgins & Taylor; OUP 2004 *Detailed practical & research aspects of Bio-Info*
"Oxford Dictionary of Biochemistry and Molecular Biology" OUP 2004 *Good for Definitions*
"Bioinformatics" Lesk; OUP 2002 *Good for placing the study of Bio-Info in a proper context*
"Organic Chemistry of Biological, McMurry & Begley; Scion 2005 *Good for seeing some curly arrows (about time); concentrates mainly on biosynthesis.*
"Biochemistry" Berg, Tymoczko & Stryer 5ed; Freeman 2002 *A great all-rounder*
"Essential Genetics, A Genomics Perspective" Hartl & Jones 4ed; Jones & Bartlett 2006

Hand-outs and **Slot-ins** will be distributed throughout the course or can be downloaded from the website (see above). These will summarize key parts and learning outcomes from the course. Collect the set.

Hand-outs

- ☐ **(F) Foundation.**
- ☐ **(1) Biological Catalysis.**
- ☐ **(2) From Genome to Proteome.**
- ☐ **(3) Chemical Methods for the Proteome.**
- ☐ **(4) Case Studies in Chemical Biology.**

Slot-ins

- ☐ **Course Summary (S)**
- ☐ **Definitions (D)**
- **Techniques (T)**
  - ☐ **T1**: DNA Seqeuncing
  - ☐ **T2**: Protein Sequencing
- **Mechanisms (M)**
  - ☐ **M1**: PCR
  - ☐ **M2**: Acyl Transfer
  - ☐ **M3**: Glycosidases
- **Case Histories (CH)**
  - ☐ **CH1**: Proteome Ser Protease Tagging
  - ☐ **CH2**: Mapping the Role of PTMs
  - ☐ **CH3**: Sugar-Processing Enzymes

Chemical Biology (& Biological Chemistry)

Ben Davis [Ben.Davis@chem.ox.ac.uk] – 4 Lectures as Part of Option II

Prior Knowledge Required

1. *1st year Biological Chemistry course*: Molecules of life. What are amino acids, peptides, proteins, sugars, nucleotides, nucleic acids – knowledge of their structures. Energy & Phosphates. Protein structure and folding. Basic principles of enzyme catalysis and inhibitor action; Michaelis-Menten kinetics. Basics of protein structure. Genetics and Protein Synthesis; transcription/translation/DNA→mRNA→protein; one letter and three letter codes for amino acids.
2. *2nd year Biological Chemistry course*: the mechanisms and principles behind primary metabolism; nature's reagents; glycolysis; CAC; oxidative phosphorylation; amino acid metablolism.

There will be one 'foundation' class (**F**) revising some of the key aspects of your 1st and 2nd year lectures as part of this course: Monday of 5th week, 5-6pm, Wolfson Seminar Room, CRL.

Books:

"Structure and Mechanism in Protein Science" Fersht; Freeman 2000 *Good on Enzyme Basics & Folding*
"An Introduction to Biotransformations in Organic Chemistry" Hanson; *A good intro to biotrans*
"Biotransformations in Organic Chemistry" Faber; *More detailed aspects of biotrans*
"Biochemistry and Molecular Biology" Elliott & Elliott; OUP *Good for quick overview & basics*
"Foundations of Chemical Biology" OCP 98 *Good as a basic introduction (e.g. folding)*
"Carbohydrate Chemistry" OCP 99 *A Masterpiece – clearly.*
"Bioinformatics" Higgins & Taylor; OUP 2004 *Detailed practical & research aspects of Bio-Info*
"Oxford Dictionary of Biochemistry and Molecular Biology" OUP 2004 *Good for Definitions*
"Bioinformatics" Lesk; OUP 2002 *Good for placing the study of Bio-Info in a proper context*
"Organic Chemistry of Biological, McMurry & Begley; Scion 2005 *Good for seeing some curly arrows (about time); concentrates mainly on biosynthesis.*
"Biochemistry" Berg, Tymoczko & Stryer 5ed; Freeman 2002 *A great all-rounder*
"Essential Genetics, A Genomics Perspective" Hartl & Jones 4ed; Jones & Bartlett 2006

Hand-outs and **Slot-ins** will be distributed throughout the course or can be downloaded from the website (see above). These will summarize key parts and learning outcomes from the

"An Introduction to Biotransformations in Organic Chemistry" Hanson; *A good intro to biotrans*
"Biotransformations in Organic Chemistry" Faber; *More detailed aspects of biotrans*
"Biochemistry and Molecular Biology" Elliott & Elliott; OUP *Good for quick overview & basics*
"Foundations of Chemical Biology" OCP 98 *Good as a basic introduction (e.g. folding)*
"Carbohydrate Chemistry" OCP 99 *A Masterpiece – clearly.*
"Bioinformatics" Higgins & Taylor; OUP 2004 *Detailed practical & research aspects of Bio-Info*
"Oxford Dictionary of Biochemistry and Molecular Biology" OUP 2004 *Good for Definitions*
"Bioinformatics" Lesk; OUP 2002 *Good for placing the study of Bio-Info in a proper context*
"Organic Chemistry of Biological, McMurry & Begley; Scion 2005 *Good for seeing some curly arrows (about time); concentrates mainly on biosynthesis.*
"Biochemistry" Berg, Tymoczko & Stryer 5ed; Freeman 2002 *A great all-rounder*
"Essential Genetics, A Genomics Perspective" Hartl & Jones 4ed; Jones & Bartlett 2006

Hand-outs and **Slot-ins** will be distributed throughout the course or can be downloaded from the website (see above). These will summarize key parts and learning outcomes from the course. Collect the set.

Hand-outs

- ☐ **(F) Foundation.**
- ☐ **(1) Biological Catalysis.**
- ☐ **(2) From Genome to Proteome.**
- ☐ **(3) Chemical Methods for the Proteome.**
- ☐ **(4) Case Studies in Chemical Biology.**

Slot-ins

- ☐ **Course Summary (S)**
- ☐ **Definitions (D)**
- **Techniques (T)**
  - ☐ **T1**: DNA Seqeuncing
  - ☐ **T2**: Protein Sequencing
- **Mechanisms (M)**
  - ☐ **M1**: PCR
  - ☐ **M2**: Acyl Transfer
  - ☐ **M3**: Glycosidases
- **Case Histories (CH)**
  - ☐ **CH1**: Proteome Ser Protease Tagging
  - ☐ **CH2**: Mapping the Role of PTMs
  - ☐ **CH3**: Sugar-Processing Enzymes

Topics to be Covered

(F) Foundations of Chemical Biology & Biochemistry. Overview of Genomic Basics; Genome; Genes. DNA sequencing (Sanger dideoxy **T1**). Transcription; Codons. Peptide & Protein Structure; Protein Folding & (in brief) Archictecture. Enzymes; Michaelis-Menten kinetics; inhibition modes; the serine protease mechanism.

D1, T1, S

(1) Biological Catalysis. General principles of enzyme catalysis. Overview of biological catalysts (including ribozymes) and classifications. General themes of catalysis that allow insight into other biological mechanisms. Catalytic antibodies. Acyl transfer as an illustration of dominant modes of catalysis (proteasome; proteases; ribosome; protein splicing and ligation; carbohydrate-processing enzymes).

D1, S, M2, M3

(2) From Genome to Proteome. Acyl Transfer in Protein Synthesis. The Mechanism of the Ribosome. Proteomes. Functional Genomics & Chemical Proteomics. Using Genomic Information. Medicinal and other implications of Genomics

(3) Chemical Methods for the Proteome. Characterisation of Proteins. CD, GE (& in brief) sequencing of peptides. Genetic Engineering (as chemistry) including recombinant methods, mutants and the polymerase chain reaction (PCR). Extent, roles & effects of Post-translational Modifications (PTMs).

T2, M1

(**4) Case Studies in Chemical Biology.** Mechanism-based probes and inhibitors. Chemical strategies for elucidating biological mechanism.

CH1, CH2, CH3

Single-letter and three-letter symbols for amino acids

Amino acid	One letter	Three letter
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val
Unspecified or unknown	X	Xaa

The genetic code Remember **U** in mRNA and **T** in DNA

5′ base	Middle base				3′ base
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA Leu	UCA Ser	UAA Stop*	UGA Stop*	**A**
	UUG Leu	UCG Ser	UAG Stop*	UGG Trp	**G**
C	CUU Leu	CCU Pro	CAU His	CGU Arg	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	**A**
	AUG Met†	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA Glu	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

*Stop codons have no amino acids assigned to them.

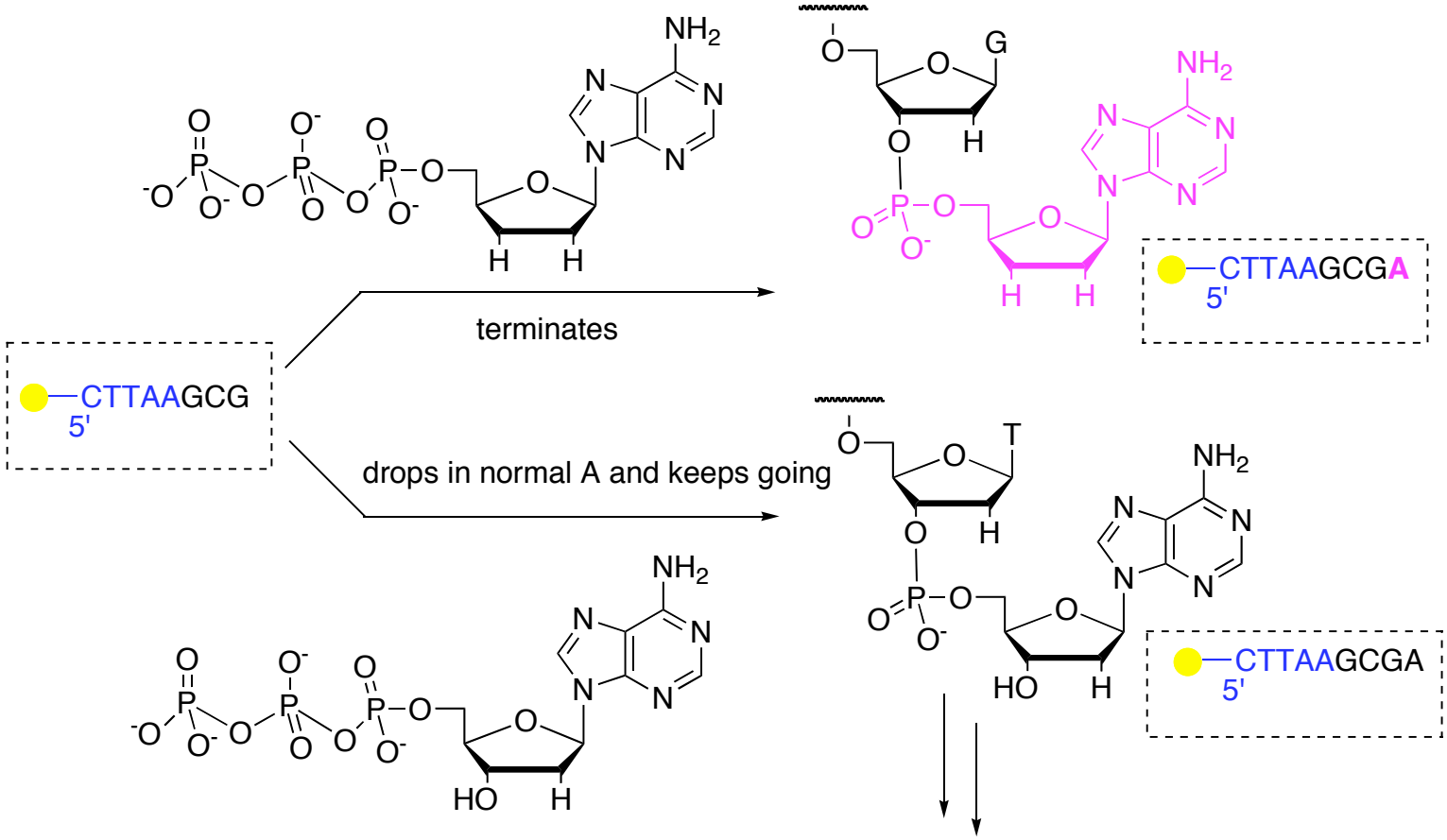
†The AUG codon is the usual initiation codon as well as that for methionine residues elsewhere. The code is almost universal but differences have been found in mitochondrial DNA from some organisms

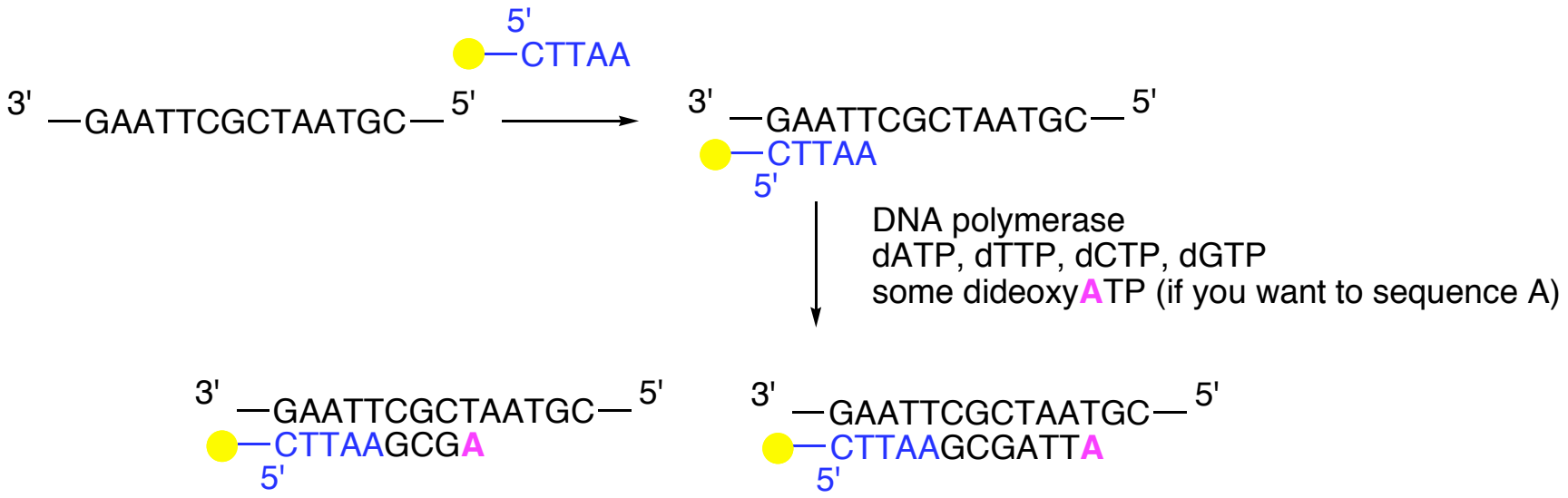
Genome sizes

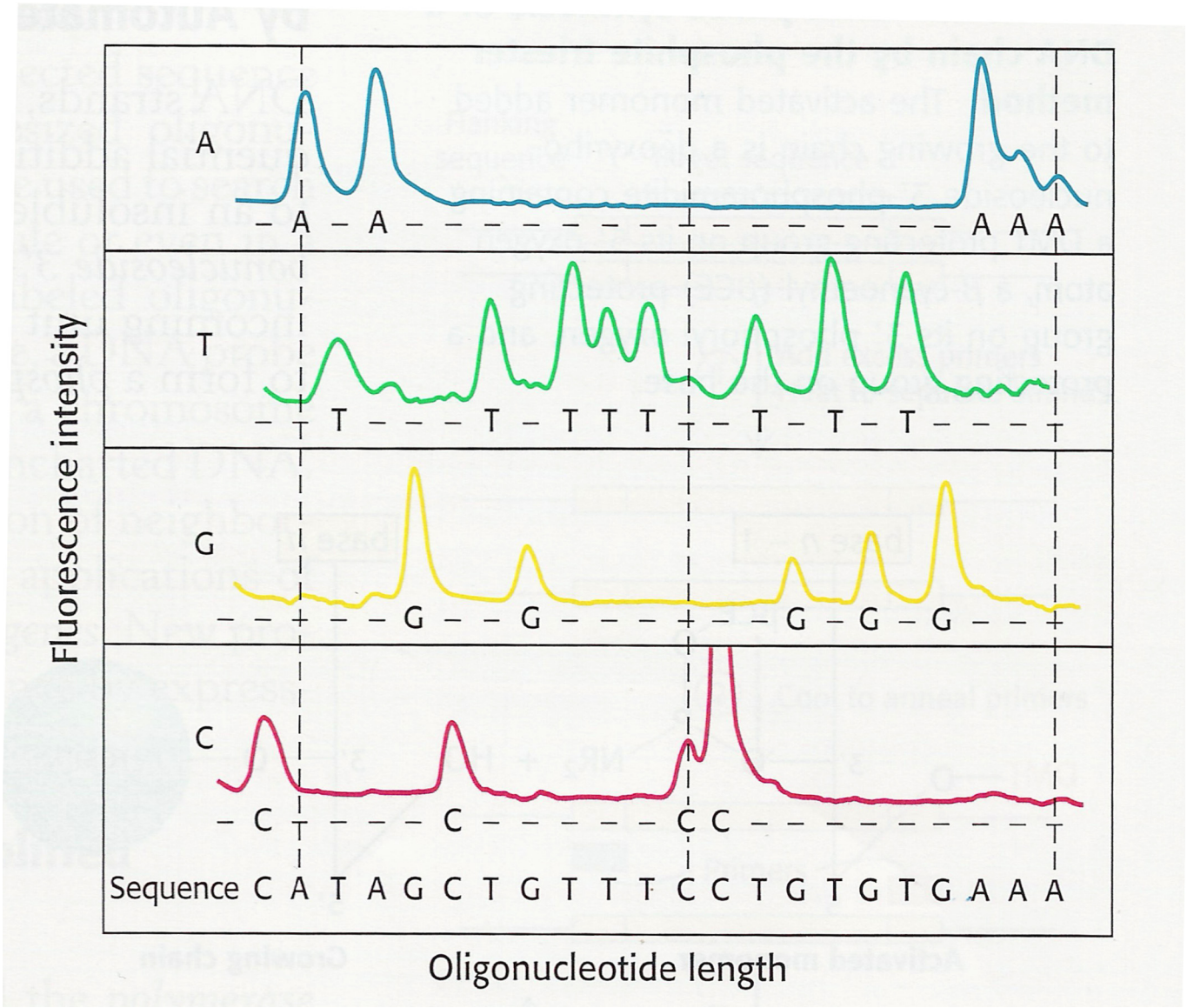
Organism	*Number of Base pairs*	*Number of Genes*	*Comment*
ϕX–174	5 386	10	virus infecting *E. coli*
Human mitochondrion	16 569	37	subcellular organelle
Epstein–Barr virus (EBV)	172 282	80	cause of mononucleosis
Mycoplasma pneumoniae	816 394	680	cause of cyclic pneumonia epidemics
Rickettsia prowazekii	1 111 523	878	bacterium, cause of epidemic typhus
Treponema pallidum	1 138 011	1 039	bacterium, cause of syphilis
Borrelia burgdorferi	1 471 725	1 738	bacterium, cause of Lyme disease
Aquifex aeolicus	1 551 335	1 749	bacterium from hot spring
Thermoplasma acidophilum	1 564 905	1 509	archaeal prokaryote, lacks cell wall
Campylobacter jejuni	1 641 481	1 708	frequent cause of food poisoning
Helicobacter pylori	1 667 867	1 589	chief cause of stomach ulcers
Methanococcus jannaschii	1 664 970	1 783	archaeal prokaryote, thermophile
Hemophilus influenzae	1 830 138	1 738	bacterium, cause of middle ear infections
Thermotoga maritima	1 860 725	1 879	marine bacterium
Archaeoglobus fulgidus	2 178 400	2 437	another archaeon
Deinococcus radiodurans	3 284 156	3 187	radiation-resistant bacterium
Synechocystis	3 573 470	4 003	cyanobacterium, 'blue-green alga'
Vibrio cholerae	4 033 460	3 890	cause of cholera
Mycobacterium tuberculosis	4 411 529	4 275	cause of tuberculosis
Bacillus subtilis	4 214 814	4 779	popular in molecular biology
Escherichia coli	4 639 221	4 406	molecular biologists' all-time favourite
Pseudomonas aeruginosa	6 264 403	5 570	largest prokaryote sequenced as yet
Saccharomyces cerevisiae	12.1×10^6	5 885	yeast, first eukaryotic genome sequenced
Caenorhabditis elegans	95.5×10^6	19 099	the worm
Arabidopsis thaliana	1.17×10^8	25 498	flowering plant (angiosperm)
Drosophila melanogaster	1.8×10^8	13 601	the fruit fly
Fugu rubripes	3.9×10^8	30 000	puffer fish (fugu fish)
Human	3.2×10^9	34 000?	
Wheat	16×10^9	30 000	
Salamander	10^{11}	?	
Psilotum nudum	10^{11}	?	whisk fern – a simple plant

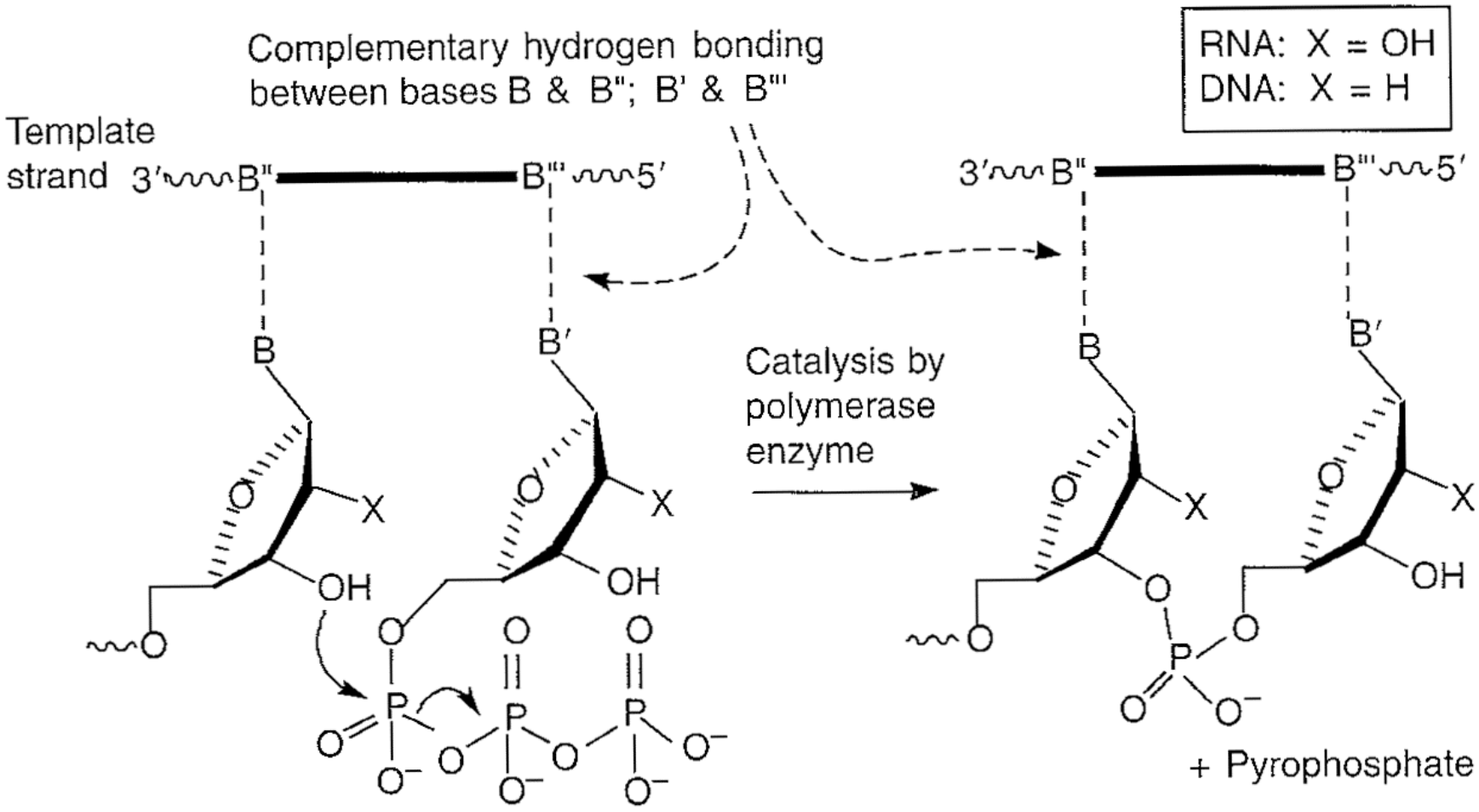
Chemical Biology	The application of the principles of chemistry to the understanding of the function of biological molecules in their cellular environment.
Biological Chemistry	The study of the biological aspects of chemistry. In this course, for example, the use of biomolecules or the strategies that occur in biological environments to achieve and enable synthetic transformations. Often also used to refer to the 'curly-arrow' mechanisms (**M**) found in biological systems.
Genome	The whole of the genetic information of an organism. DNA in prokaryotes and eukaryotes. DNA and/or RNA in viruses. Only 1 per organism (despite some older definitions).
Eukaryote	Any organism whose cells contain a nucleus bounded by a nuclear membrane and containing true chromosomes. It is a characteristic of all organisms except bacteria, archea and cyanobacteria. They undergo meiosis.
Prokaryote	An organism in which the genomic DNA is not enclosed in a nuclear membrane within the cells.
Expression	Really *Gene Expression*. The process by which genes become manifest as a phenotype. This normally means DNA→mRNA→protein = *transcription*+*translation*.
Transcription	The synthesis of RNA on a template of DNA. This is done by mRNA polymerases.
Introns	DNA sequence that interrupts the coding-sequence of a gene that is cut out post-transcription.
Exons	Part of a gene whose sequence is present in mature mRNA post-splicing.
Translation	The synthesis of polypeptide as determined by the sequence and composition of mRNA.This takes place in the ribosome and the 'coding' is achieved through the use of 64 permutations of triplet *codons*. These denote, start, stop or one of 20 naturally-occurring amino acids.

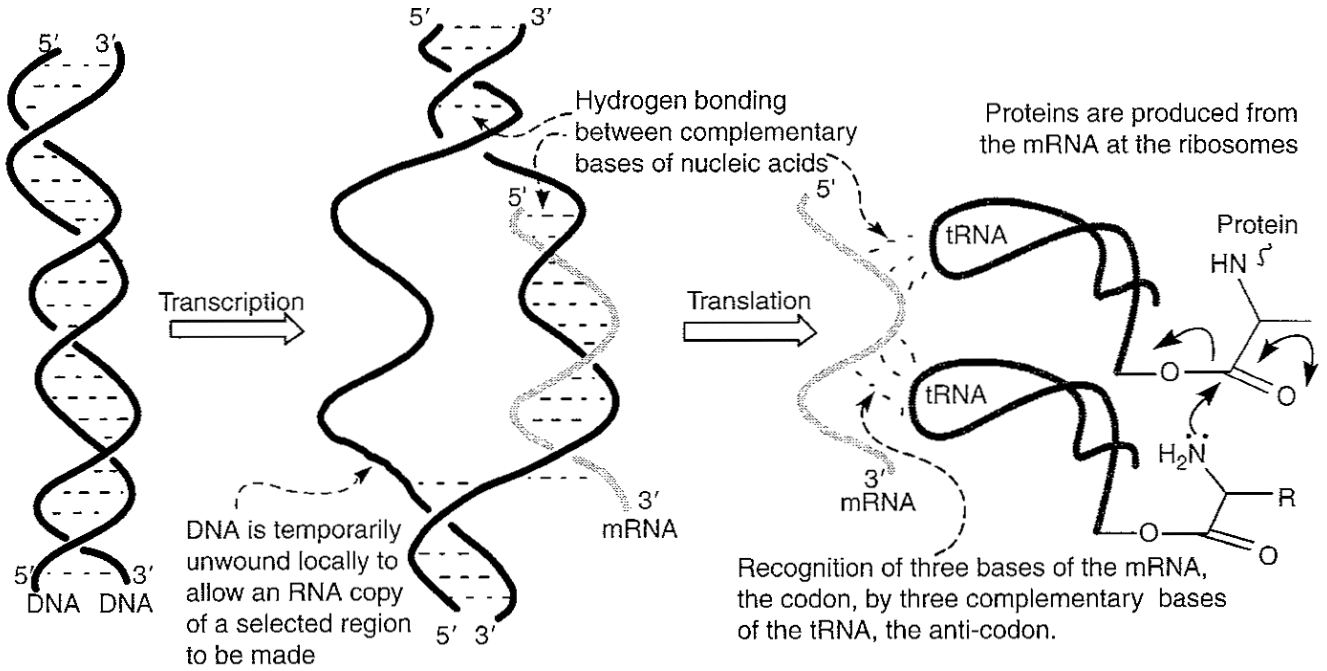
	It is a characteristic of all organisms except bacteria, archea and cyanobacteria. They undergo meiosis.
Prokaryote	An organism in which the genomic DNA is not enclosed in a nuclear membrane within the cells.
Expression	Really *Gene Expression*. The process by which genes become manifest as a phenotype. This normally means DNA→mRNA→protein = *transcription*+*translation*.
Transcription	The synthesis of RNA on a template of DNA. This is done by mRNA polymerases.
Introns	DNA sequence that interrupts the coding-sequence of a gene that is cut out post-transcription.
Exons	Part of a gene whose sequence is present in mature mRNA post-splicing.
Translation	The synthesis of polypeptide as determined by the sequence and composition of mRNA.This takes place in the ribosome and the 'coding' is achieved through the use of 64 permutations of triplet *codons*. These denote, start, stop or one of 20 naturally-occurring amino acids.

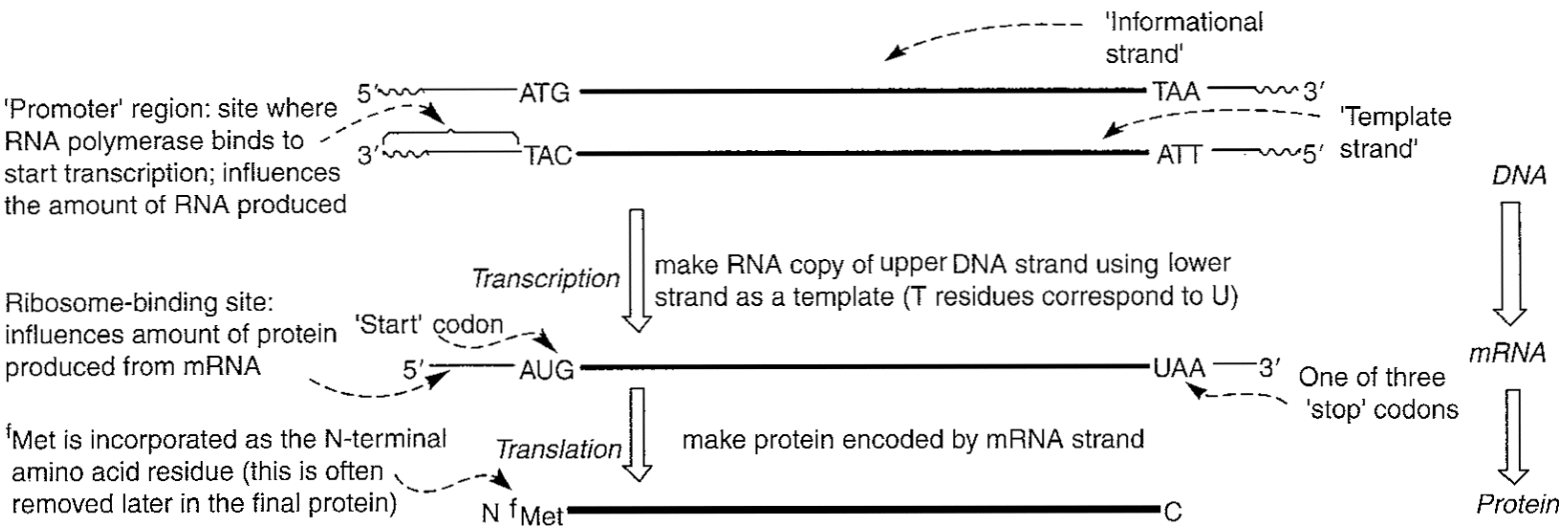

terminates
drops in normal A and keeps going
CTTAAGCG
5'
CTTAAGCGA
5'
CTTAAGCGA
5'


5'
CTTAA
3'
GAATTCGCTAATGC
5'
3'
GAATTCGCTAATGC
5'
CTTAA
5'
DNA polymerase
dATP, dTTP, dCTP, dGTP
some dideoxyATP (if you want to sequence A)
3'
GAATTCGCTAATGC
5'
CTTAAGCGA
5'
3'
GAATTCGCTAATGC
5'
CTTAAGCGATTA
5'


A
T
G
C
Fluorescence intensity
– A – A – – – – – – – – – – – – – – A A A
– – T – – – T – T T T – – T – T – T – – – –
– – – – – G – – G – – – – – – G – G – G – – –
– C – – – – C – – – – – – C C – – – – – – – –
Sequence C A T A G C T G T T T C C T G T G T G A A A
Oligonucleotide length


Complementary hydrogen bonding
between bases B & B"; B' & B"'
RNA: X = OH
DNA: X = H
Template
strand
3'
B"
B"'
5'
B
B'
X
OH
O
P
O⁻
Catalysis by
polymerase
enzyme
+ Pyrophosphate


5′ 3′
5′ 3′
DNA DNA
Transcription
5′ 3′
Hydrogen bonding
between complementary
bases of nucleic acids
5′
3′
mRNA
DNA is temporarily
unwound locally to
allow an RNA copy
of a selected region
to be made
5′ 3′
Translation
Proteins are produced from
the mRNA at the ribosomes
5′
tRNA
tRNA
Protein
HN
O
O
H_2N
R
O
O
3′
mRNA
Recognition of three bases of the mRNA,
the codon, by three complementary bases
of the tRNA, the anti-codon.

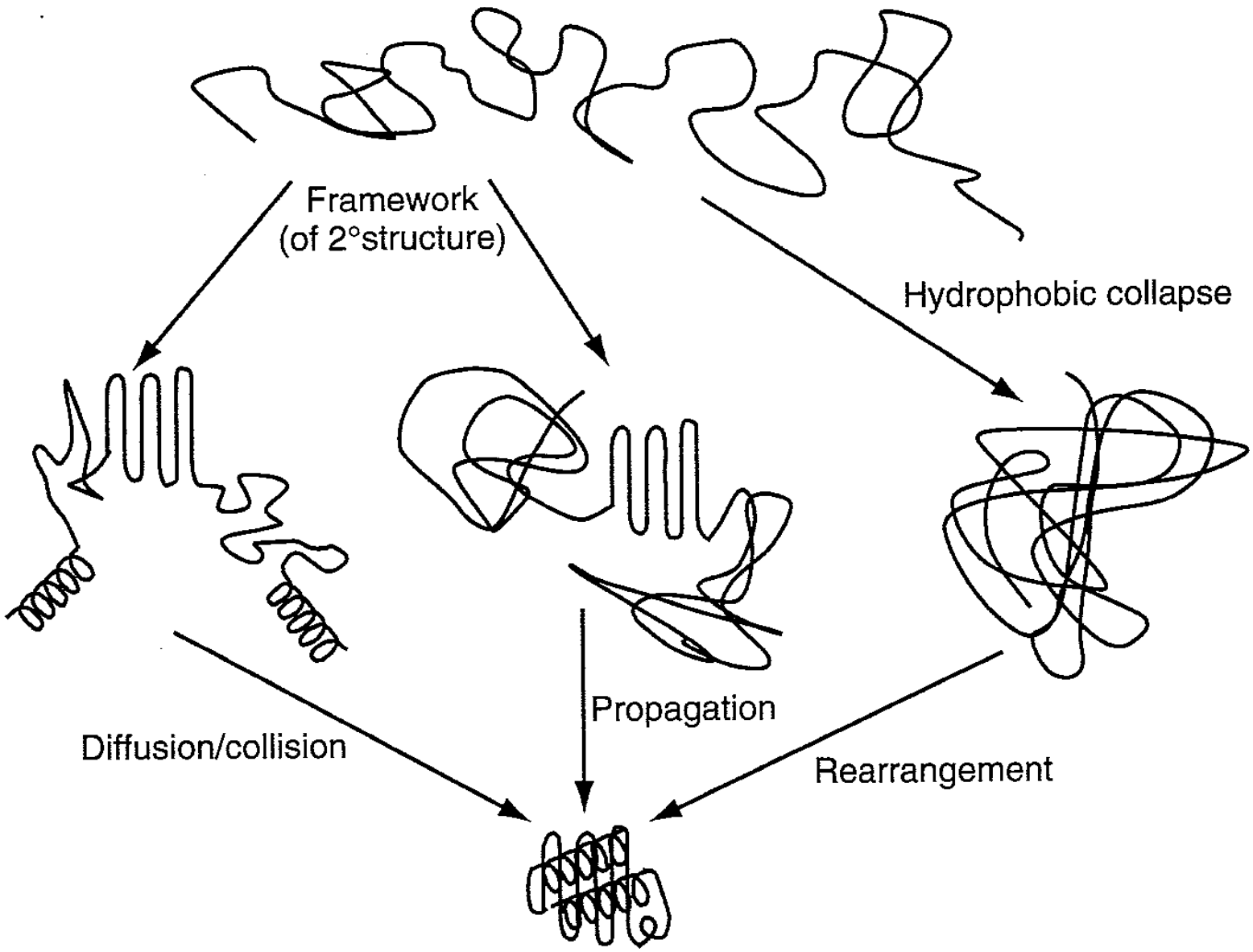

'Informational
strand'
'Promoter' region: site where
RNA polymerase binds to
start transcription; influences
the amount of RNA produced
5′ ATG TAA 3′
3′ TAC ATT 5′
'Template
strand'
DNA
Transcription
make RNA copy of upper DNA strand using lower
strand as a template (T residues correspond to U)
Ribosome-binding site:
influences amount of protein
produced from mRNA
'Start' codon
5′ AUG UAA 3′
One of three
'stop' codons
mRNA
fMet is incorporated as the N-terminal
amino acid residue (this is often
removed later in the final protein)
Translation
make protein encoded by mRNA strand
N fMet C
Protein

The genetic code Remember **U** in mRNA and **T** in DNA

5′ base	Middle base				3′ base
	U	**C**	**A**	**G**	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA Leu	UCA Ser	UAA Stop*	UGA Stop*	**A**
	UUG Leu	UCG Ser	UAG Stop*	UGG Trp	**G**
C	CUU Leu	CCU Pro	CAU His	CGU Arg	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	**A**
	AUG Met†	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA Glu	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

*Stop codons have no amino acids assigned to them.

†The AUG codon is the usual initiation codon as well as that for methionine residues elsewhere. The code is almost universal but differences have been found in mitochondrial DNA from some organisms


Framework
(of 2°structure)
Hydrophobic collapse
Propagation
Diffusion/collision
Rearrangement

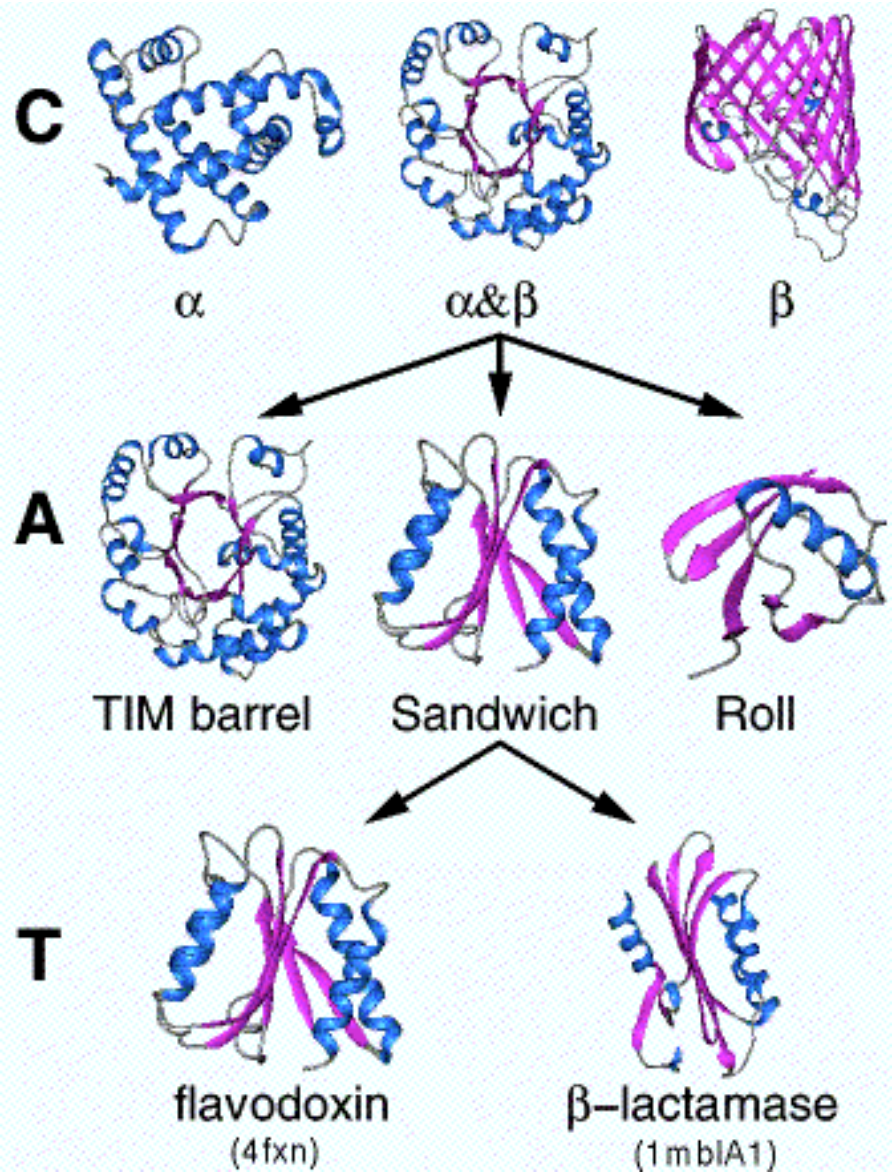
A database for 3D shapes of protein domains

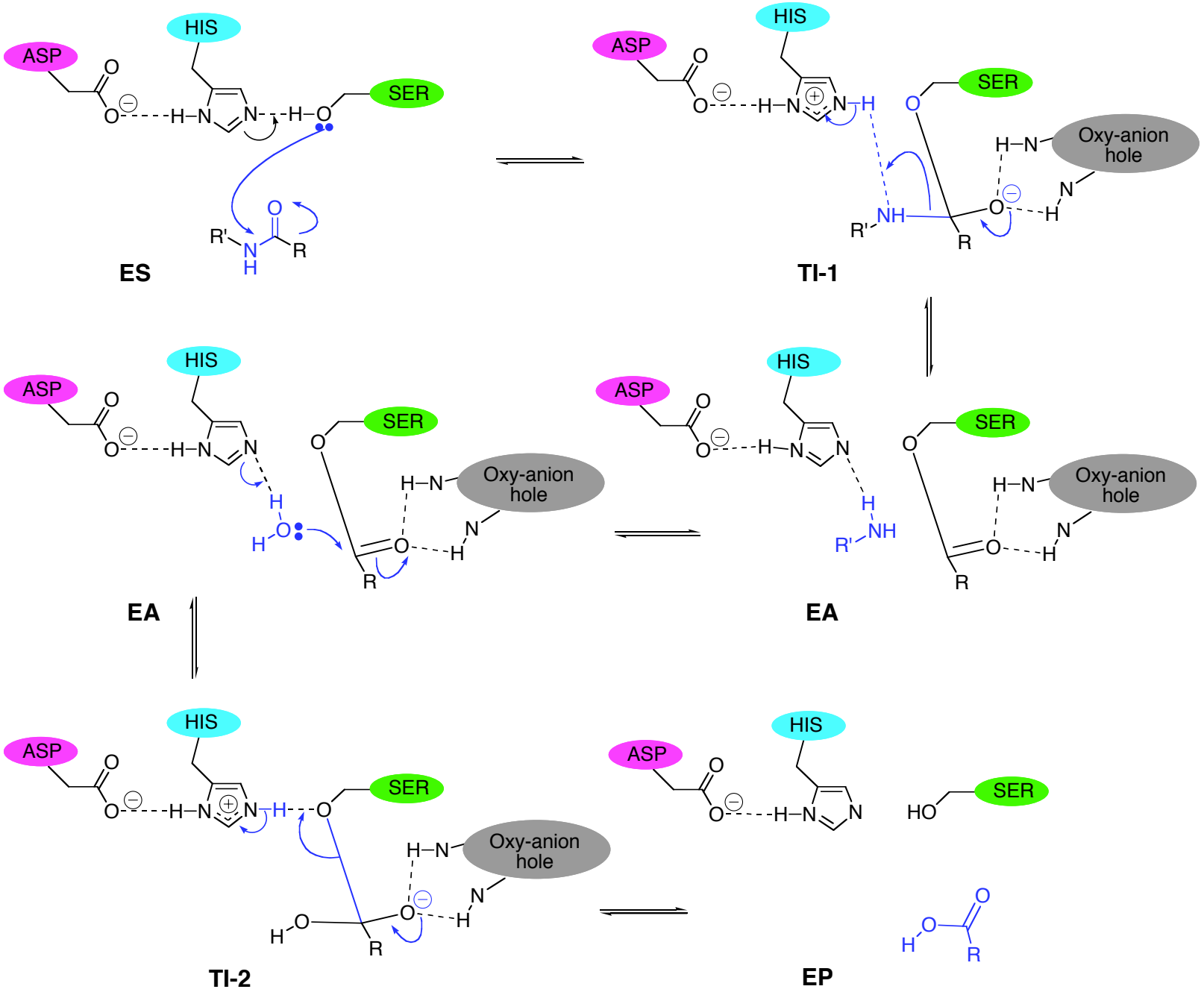
C Class – Secondary Structure

A Architecture – Arrangements of 2ndary Structure but ignores connections

T Topology – Clustering of A

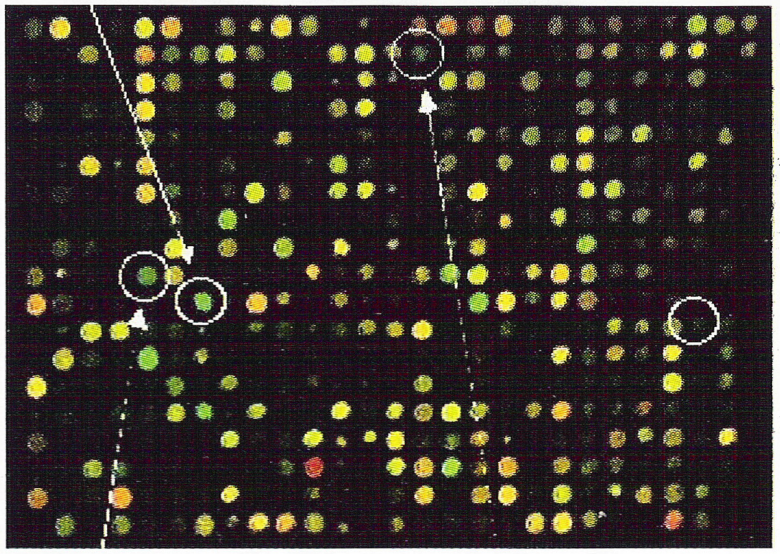
H Homologous Superfamliy – Combines Sequence and Structure information to create groups.

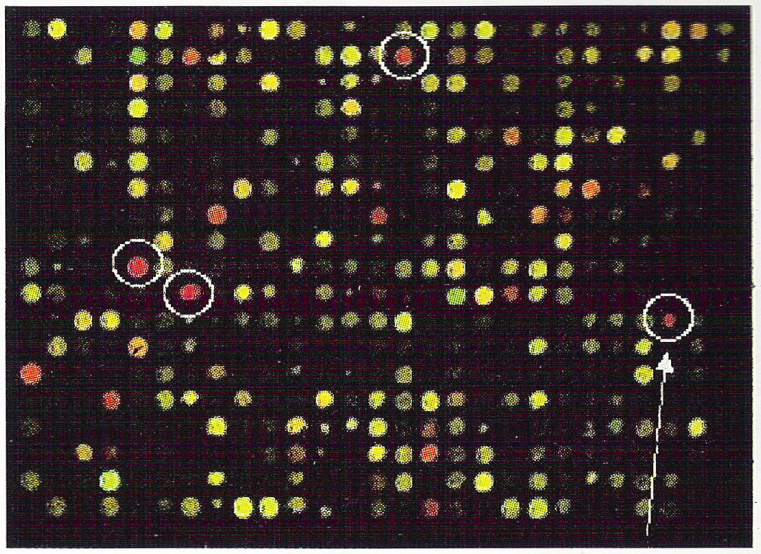



ASP
HIS
SER
Oxy-anion hole
ES
TI-1
EA
EA
TI-2
EP

Lysyl oxidase uninfected

8 wk post infected





Collagen binding protein Procollagen, type III, alpha 2

Procollagen, type V, alpha 2